VirusX: simulation of viral conformational changes

Yinglong Miao

Advisor: Peter J. Ortoleva

Center for Cell and Virus Theory, Department of Chemistry, Indiana University,

Bloomington, Indiana 47405

Contents

1. Objectives 3

2. Background 3

3. Methods 10

3.1 PDB2PQRM 10

3.2 Icosahedral Virus Capsid Generator 12

3.3 3D nonlinear PB model 13

3.4 Space-warping Method 17

4. Results and Discussion 19

4.1 Icosahedral Virus Capsid Generator 19

4.2 PB calculations 20

5. Conclusion 27

Acknowledgements 28

References 28

1. Objectives

Up to April 12, 2005, there are 30,453 structures solved by X-ray and/or NMR in Online Protein Data Bank[1] (http://www.pdb.org/) and 1,746 of them are viral structures. What can we do with so many available structures of viruses? Methods and modeling are needed to investigate their behaviors in bio-electrolyte medium. The aim of this project is to develop a large scale and general virus simulator, VirusX, to simulate viral conformational changes in electrolyte medium on the basis of Molecular Dynamics — space-warping method[2] and a Poisson-Boltzmann model[3]. It integrates efficient all-atom computations using generalized coordinates and a continuum approach to describe the electrolyte medium and account for the channeling of the electric field along a macromolecule.

2. Background

It is known that the morphology and stability of many viruses are affected by the conditions in the host medium. In particular pH and ionic strength are known to affect the morphology of many viruses such as tomato bushy stunt virus (TBSV) [4], turnip crinkle virus[5], and cowpea chlorotic mottle virus (CCMV) [6, 7]. In this project, VirusX will be developed to study these conformational changes, especially the self-assembly of virus capsids.

Viruses are made of large components (e.g. protein, RNA or DNA) with a huge

number of atoms. A typical virus capsid contains more than 100,000 atoms. For example, a CCMV protein capsid contains (PDB ID: 1cwp) about 430,000, a Human Rhinovirus (HRV, PDB ID: 1r1a) capsid contains about 380,000 atoms, and a Human Hepatitis B virus (HBV, PDB ID: 1qgt) capsid contains 543,480 atoms. With current computation power, it is not feasible to do all-atom calculations using conventional Molecular Dynamics (MD). To solve this problem, viral structures input to VirusX will be divided into fragments (such as chains, protomers, pentamers and hexamers in its capsid, etc.) according to their structural patterns. The electrostatic and Van der Waals interactions between fragments will be calculated and a set of "global coordinates" will be applied on them to simulate the viral conformation changes under the framework of MD. However, VirusX will keep an option capable of implementing all atomic-scale simulations or multiple scale calculations, a set of "global coordinates" to capture the viral overall behavior and a set of "atomic coordinates" to keep the atomic scale details. When the computational power becomes available, such calculations can be implemented.

In calculating electrostatic interactions, it is important to account for the channeling of the electric field along a macromolecule due to the dielectric constant contrast between the high-dielectric medium and the low-dielectric interior of a protein, an effect not accounted for in 1/r Coulomb computations used in the force fields, like CHARMM[8]. The electrostatic interactions between two charges in a protein may be larger through the interior protein channel than the 1/r Coulomb

computations through the high-dielectric medium, which will be modified with our 3D nonlinear Poisson-Boltzmann (PB) model[2].

Electrostatic interactions can be defined as forces and energies resulting from an asymmetric charge distribution or a non-zero net charge of a solute. They play a key role in the structure and function of macromolecules, especially biomolecules. DNA is one typical example of them with negative charges due to phosphate groups. However, there are two main challenges in this area, long-range interactions and electronic polarizability. It is computationally expensive to simulate them accurately. Among current simulation methods, the PB equation is a commonly used model, which treats these two aspects with a continuum field variable, the dielectric constant.

In PB model for macromolecular electrostatics, the macromolecules (solute), which contain atomic charges, are treated as a low-dielectric region bounded by the molecular surface. The atomic charges are typically drawn from a molecular mechanics force field. Around the molecule is the high-dielectric solvent, which may contain mobile ions caused by dissolved electrolytes. The PB model accounts for long-range interactions fully and electronic polarizability implicitly with the dielectric constants of the aqueous media.

As a prerequisite, the atomic charges inside a macromolecule must be drawn from an empirical force field before implementing PB model. Following the first force field,

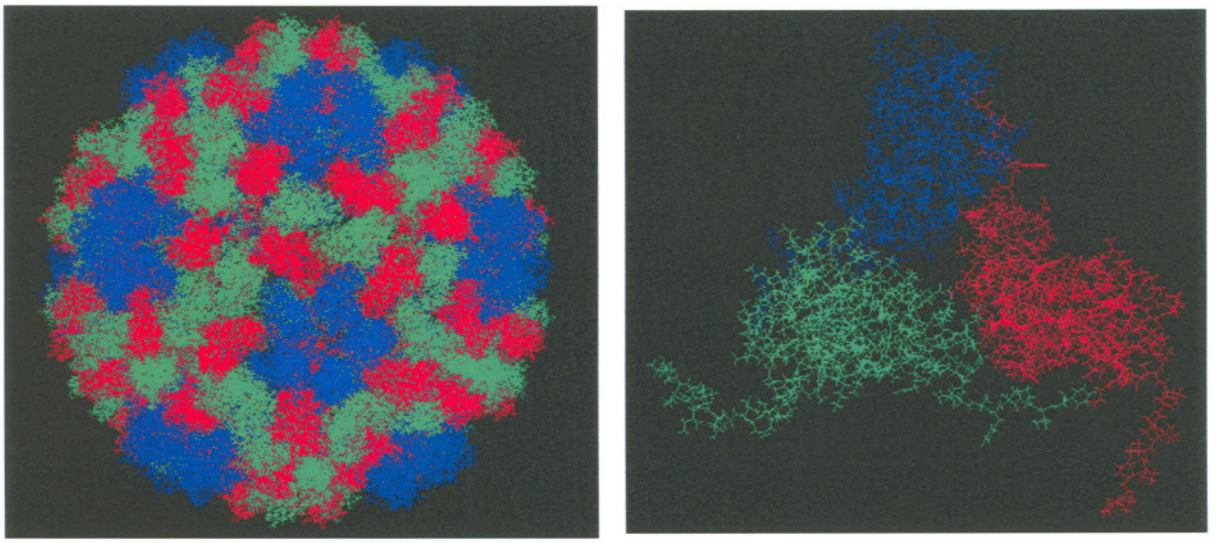
CHARMM, many other empirical force fields, such as AMBER, GROMOS, OPLS, etc. have been developed to simulate biomolecules[9]. Terms in the energy functions are modified to make force fields more accurate. For a given macromolecule, using different force fields will lead to different results. Before implementing a force field it is important to be aware of the approaches used in its development first. For a certain class of biomolecules, some force fields may generate better results than others. Accordingly, force fields are also categorized as protein force fields, nucleic acid force fields, lipid force fields, force fields for carbohydrates, and so on. With continuing advances in computational hardware, feasible implementation of quantum mechanical calculation in biomolecules will generate accurate force fields.

After obtaining atomic charges inside the macromolecule, we need to study the medium around it. The PB model accounts for solvent molecules implicitly via the dielectric constant, which is low within the molecule and is assumed to increase gradually to the unperturbed, far-field value over several angstroms away from the molecule of interest[10]. In our group's previous research work, *Sayyed-Ahmad et al.*[3] used an exponential function to smooth such transitional variation of dielectric constant in the molecule-solvent interface. The Poisson Equation accounts for mobile ions in the solvent via a highly nonlinear term due to the exponential dependence of the mobile ion concentrations on the potential. This nonlinear and elliptic PB equation is reformulated to an advection-diffusion equation, which is solved with 3D alternative direction implicit (ADI) method and finite difference method efficiently.

This 3D nonlinear PB solver will be integrated into VirusX to the correct 1/r Coulomb computations used in the force fields.

To test VirusX, the pH and ionic strength dependence and swollen behavior of CCMV will be investigated. CCMV is a member of the bromovirus group of the Bromoviridae family. The genome of CCMV consists of four positive-sense single-stranded RNA molecules encapsulated in three virions. Because the purified RNA and coat protein of CCMV can reassemble in vitro to produce infectious virions, CCMV is an excellent system for studying protein-protein and protein-RNA interactions, which provide important information in the assembly and disassembly of icosahedral viruses.

The crystal structure of CCMV was solved at 3.2Å resolution by x-ray crystallography[6]. The CCMV capsid (Figure 1a) is made up of 180 chemically identical protein subunits that form a 286-Å-diameter icosahedral shell display T=3 quasisymmetry. Each protein subunit (one branch of the three quasi-threefold related subunits shown in Figure 1b) is composed of 190 amino acids taking three quasi-equivalent positions on the virus surface.



(a) (b)

Figure 1 (a) The CCMV capsid (a T=3 crystal lattice for a truncated icosahedron) and (b) three quasi-threefold related subunits in CCMVV protomer (A in blue, B in red and C in green) (Images prepared with MDL Chime).

It has been mentioned that the morphology and stability of CCMV was found to be affected the conditions in host medium change, such as pH, temperature and ionic strength[6, 7]. Native CCMV is stable around pH 5.0 in a compact form. When the pH of the solution is raised to 7.0 at low ionic strength (<0.2M) in the absence of divalent cations, the CCMV capsid undergoes a 10% radial expansion at the quasi-threefold axes. In the expansion scheme[6] proposed by *H. Liu et al.*, the internal structures of pentamers and hexamers in the CCMV capsid are fixed as rigid units and the expansion was simulated as a 1.2-Å translation and a 0.45° counterclockwise rotation along the fivefold axis for the pentamers, and a 1.05-Å translation and a 0.4° counterclockwise rotation along the icosahedral threefold axis for the hexamers. The mechanism and pathway of this pH induced swelling process is also studied by *Tama*

et al. with normal mode analysis[12].

Apart from studies on CCMV capsid, *Zhang et al.*[13] simulated electrostatic interaction between RNA and proten capsid in CCMV with a coarse-grain RNA model and a Monte Carlo approach. In the coarse-grain RNA model, each RNA nucleotide was treated as an isolated sphere with a charge of -0.25e and no connection was enforced between RNA spheres. Results show that there is a very strong interaction between the N-terminal residues of the capsid proteins, which are highly positive charged, and the viral RNA. The distribution of RNA nucleotides inside CCMV capsid was found to form a shell close to the capsid with the highest densities associated with the capsid dimers. These high-density regions are connected to each other in the shape of a continuous net of triangles. Medium density of RNA is found under the pentamers of the capsid.

There was also a theoretical attempt by *Isea et al.*[14] to study the RNA releasing process of CCMV. In the model, a hypothetical, short and single stranded RNA, [5'-R(PGpGpApCpUpUpCpGpGpUpCpC)-3')], was constructed and only the three quasi-threefold related subunits (a protomer) of CCMV capsid was included in the calculations. The molecular dynamics simulations indicate that the RNA losses its secondary structure and movies into the channel on the three quasi-threefold axis of CCMV capsid by free diffusion.

In the simulation test of CCMV capsid with VirusX, the internal structures of pentamers and hexamers in CCMV capsid will be fixed and assumed as movement units during the swelling. Electrostatic interactions and Van der Waals interactions between such rigid units will be calculated to obtain the energy-minimized structure of CCMV capsid under different pH at low ionic structure (0.15M). If allowed by computer power, the RNA inside the CCMV capsid and the cell surface will be included in the simulations as a complete system.

3. Methods

3.1 PDB2PQRM

In order to investigate viral conformational changes, enough information about viral molecules including their structural data and force field parameters, such as atomic charges, radii and masses, must be known. With PDB structures downloaded, their structural information is included and a program is needed to look up atomic charges, radii and masses from a force field.

Based on Tinker (http://dasher.wustl.edu/tinker/) version 4.1, I developed PDB2PQRM, a program used to assign force field parameters for atoms in PDB structures. It converts PDB files to PQRM files, in which the PDB file format and its header information of biomolecular structures are kept, while the occupancy and temperature factor for every atom are replaced with its charge and radius, and the

atomic mass is added to the end. PQRM files contain all the information of atoms needed to be served as input to VirusX: atomic positions ("P"), charges ("Q"), radii ("R") and mass ("M").

In PDB2PQRM, there is one problem left to VirusX due to incomplete PDB source structures. In most PDB structures hydrogen atoms are not contained and some heavy atoms in certain residues and/or whole residues are missed. It is neither complete nor accurate to use them for theoretical simulations. Regarding missing hydrogens, Tinker 4.1 can add them to the PDB structures and reconstruct their coordinates, while for heavy atoms and missing residues, it cannot be easily done with Tinker 4.1. Upon completion of VirusX, those heavy atoms and/or whole residues missed in PDB files can be added to the structures and energy minimized to new structures. Then these complete structures can be input to VirusX to run the program again to obtain the final structures in the desired bio-electrolyte medium.

Before PDB2PQRM was developed, there had been a web server PDB2PQR[15] (http://agave.wustl.edu/pdb2pqr/), which assigns atomic charges and radii to PDB structures and prepares input for PB electrostatics calculations[17, 18]. However, i ts output PQR files do not contain the atomic masses, which are needed to run MD calculations. Besides, it discards the biomolecular Chain IDs from PDB files, which are needed to divide the virus capsid into fragments when implementing virus capsid generator to obtain input for VirusX. In the program I implemented, all these

information is preserved and formatted to VirusX input.

3.2 Icosahedral Virus Capsid Generator

A viral structure downloaded from PDB is just a symmetric unit in an icosahedral virus capsid. To do simulations on complete virus capsids, a method is needed to generate them first. In a virus PDB structure, a method is usually given to generate the corresponding virus capsid, but we need a general method for every virus structure. According to the symmetry properties of icosahedral viruses, the Virus Particle Explorer[16] (VIPER: http://mmtsb.scripps.edu/viper/) research team developed an Oligomer Generator to generate the complete virus capsids with a set of 60 matrices. Before utilizing it, PDB coordinates needed to be converted into VIPER coordinates with PDB to VIPER transformation matrices.

In Icosahedral Virus Capsid Generator (IVCG) I developed, I start from the PQRM files obtained from our PDB2PQRM program and apply matrices from VIPER to generate the complete virus capsids, which still contain the force field parameters (charges, radii and masses). Next, the output virus capsid of IVCG is divided into logical fragments (such as chains, protomers, pentamers and hexamers in its capsid, etc.). This final divided structure serves as the input for VirusX.

3.3 3D nonlinear PB model

Starting from Poisson's equation for a relationship between electrostatic potential φ and charge density ρ:

$$\vec{\nabla} \bullet \left(\varepsilon \vec{\nabla} \varphi\right) = -4\pi\rho \tag{1}$$

In equation (1), ε is the relative dielectric constant, 1 for vacuum and approximately 80 for liquid water at room temperature. The boundary conditions are:

$$\lim_{r \to \infty} \varphi(r) = 0 \tag{2}$$

Actually, in real computational simulations φ is truncated at a distance where it decreases to be negligible. To solve Poisson's equation (1) of a system, we also need to figure out its charge densities and relative dielectric constants.

Let's see how to calculate charge distribution of the system first. Consider a macromolecule immersed in an electrolyte. The atomic fractional charges q_i can be calculated from a parameterized force field (e.g., CHARMM[8]). For electrolyte, the mobile ions are account for the total charge density, which is given by:

$$\rho\left(\vec{r}, \underline{c}\right) = \rho_m + \rho_{ions} = \sum_{i=1}^{N_{charges}} q_i \delta(\vec{r} - \vec{r}_i) + F \sum_{i=1}^{N_{ions}} z_i c_i(\vec{r}) \tag{3}$$

where q_i are fixed atomic charges inside the macromolecules, $\vec{r}_i$ are their corresponding positions, F is Faraday's constant, Z_i and $C_i(\vec{r})$ are the valence and concentration of the *i*th type mobile ion at position $\vec{r}$. To obtain the concentration for each ionic species, we need to use the Debye-Huckel Theory of the equilibrium thermodynamic relation for chemical potential of a dilute electrolyte[8]:

$$\mu_i = \mu_i^{**} + RT \ln c_i(\vec{r}) + z_i F \varphi(\vec{r}) \tag{4}$$

where μ_i^{**} is the chemical potential pure solvent, T is the temperature. From (4), we can write:

$$c_i(\vec{r}) = \overline{c_i} \exp(-F z_i \varphi(\vec{r}) / RT) \tag{5}$$

If the potential $\varphi(\vec{r})$ far from the molecule is zero, the concentration of mobile species *i*, denoted as $\overline{c_i}$ will be equal to its bulk concentration. Equation (5) establishes a nonlinear relationship (Boltzmann distribution) between the electrostatic potential and the mobile ion concentrations. Now we can rewrite (3) as:

$$\rho(\vec{r}, \underline{c}) = \rho_m + \rho_{ions} = \sum_{i=1}^{N_{charges}} q_i \delta(\vec{r} - \vec{r}_i) + F \sum_{i=1}^{N_{ions}} z_i \overline{c_i} \exp(-F z_i \varphi(\vec{r}) / RT) \tag{6}$$

Substituting (6) in (1), we have the nonlinear Poisson-Boltzmann equation:

$$\vec{\nabla} \bullet (\varepsilon \vec{\nabla} \varphi) + 4\pi \sum_{i=1}^{N_{charges}} q_i \delta(\vec{r} - \vec{r}_i) + 4\pi F \sum_{i=1}^{N_{ions}} z_i \overline{c_i} \exp(-F z_i \varphi(\vec{r}) / RT) = 0 \tag{7}$$

For the dielectric constant ε, an exponential function is used to smooth the transitional variation (a step function approximation) in the interface of molecule and the medium with phenomenological approach:

$$\varepsilon(\vec{r}) = \min_{i=i.....N_{atoms}} \{\varepsilon_m + (\varepsilon_{atom} - \varepsilon_m) \exp(-\alpha(|\vec{r} - \vec{r}_i| - R_i)^n / R_i^n)\} \tag{8}$$

where ε_m and ε_{atom} are dielectric constants of the medium and molecular atoms, α and *n* are phonological constants, $\vec{r}_i$ is the position of the center of atom *i* in the molecule, and R_i is the radius of the atom. In the 3D nonlinear PB model, α and *n* are taken as 1 and 2, respectively, and the smoothing function is reduced to be a Gaussian function.

To solve the highly nonlinear PB equation (7), *Sayyed-Ahmad et al.*[3] minimized a variational functional with respect to electrostatic potential and reformulate it to an advection-diffusion equation (9) with Langevin steepest descent approach:

$$\frac{\partial \varphi}{\partial t} = \vec{\nabla}^2 \varphi + \frac{\vec{\nabla}\varepsilon \bullet \vec{\nabla}\varphi}{\varepsilon} + \frac{4\pi}{\varepsilon} \sum_{i=1}^{N_{charges}} q_i \delta(\vec{r} - \vec{r}_i) + \frac{4\pi F}{\varepsilon} \sum_{i=1}^{N_{ions}} z_i \bar{c}_i \exp(-F z_i \varphi(\vec{r}) / RT) \quad (9)$$

Comparing (9) with (7), we can find that the steady-state solution ($\frac{\partial \varphi}{\partial t}$=0) to advection-diffusion equation (9) corresponds to the solution to the elliptic equation (7), which has a unique solution. Here, we come to this advection-diffusion equation, which can be solved efficiently with 3D Douglas alternating direction method (ADI).[9]

The original code of 3D nonlinear PB solver was developed by A. Sayyed-Ahmad and K. Tuncay under Fortran 77. I rewrote it using Fortran 90 because Fortran 90 can be used to mimic Object-Oriented programming and many of its new fearures are needed to optimize our PB solver. In my Fortran 90 code of PB solver, many variables of an object are integrated into data types (e.g. define type "atom" with elements "position(3)", "mass", "charge", and "radius" to describe an atom) and related operations are encapsulated into logical modules. The code becomes clearer and easier to manipulate because of information-hidden. Besides, it becomes possible to allocate arrays dynamically during program running. It is much more efficient in memory allocation. The code is also parallelized with OpenMP to be implemented on shared-memory machines. It solves the electrostatic potential of a macromolecule in electrolyte medium very efficiently on IBM-SP computers.

Although the above 3D nonlinear PB model will be adopted through this project, it is necessary to explain the linear PB model for following discussion (Part 4.2). As a simpler model, Equation (7) is often linearized by assuming that $F\varphi(\vec{r})$ is much smaller than RT. With this assumption,

$$\exp(-Fz_i\varphi(\vec{r})/RT) \approx 1 - Fz_i\varphi(\vec{r})/RT \tag{10}$$

Considering the potential out of the molecule surface, we have:

$$\sum_{i=1}^{N_{charges}} q_i \delta(\vec{r} - \vec{r}_i) = 0 \tag{11}$$

For $\overline{c_i}$ are bulk concentrations of mobile ions far from the interface, they should follow the charge neutrality condition:

$$\sum_{i=1}^{N_{ions}} z_i \overline{c_i} = 0 \tag{12}$$

Substituting Equation (7) with the above three equations, we obtain the linear Poison-Boltzmann equation:

$$\vec{\nabla} \bullet \left(\varepsilon \vec{\nabla}\varphi\right) = \frac{8\pi F^2 I}{RT}\varphi(\vec{r}) \tag{13}$$

where I is the electrolyte ionic strength:

$$I = \sum_{i=1}^{N_{ions}} {z_i}^2 \overline{c_i} \tag{14}$$

Assuming the dielectric constant ε to be a constant all over the electrolyte media, Equation (13) can be rewritten as:

$$\vec{\nabla}^2\varphi = \varphi(\vec{r})/{\lambda_D}^2, \quad {\lambda_D}^2 \equiv \frac{\varepsilon RT}{8\pi F^2 I} \tag{15}$$

In one dimension, the solution to Equation (15) is:

$$\varphi(x) = \varphi(0)\exp(-x/\lambda_D) \tag{16}$$

As $x \to 0$, $\varphi(x) \to 0$ and the Debye length λ_D is the characteristic thickness of the region wherein potential φ is appreciable. In this project, a potential cutoff of two Debye lengths is adopted for boundary conditions (2).

3.4 Space-warping Method

Space-warping method[2] is an innovative approach to simulate large-scale conformational changes of macromolecules. It uses a set of "global coordinates" to capture the overall behavior of macromolecules and the set of "atomic coordinates" to keep the atomic scale details. Compared with simulations with atomic coordinates only, the method tends to find lower energy minima more efficiently when a macromolecule has multiple energy minima.

In the method, a set of vector functions $\vec{f}_n(\vec{r})$ (polynomial, harmonic, or wavelet) in three-dimensional space $\vec{r}$ is introduced. Consider a transformation mapping a point ($\vec{r}$) in a space to a new point ($\vec{r'}$):

$$\vec{r'} = \sum_n \Gamma_n \vec{f}_n(\vec{r}) \qquad (1)$$

The transformation flexibility depends on the choice of the number and form of the transformation functions $\vec{f}_n(\vec{r})$. The Γ_n's form a set of global coordinates that determine the actual transformation. It can capture the translation, dilation, bending and twisting of macromolecules.

When doing simulations of macromolecular conformational changes, a transformation similar to that of Equation (1) can be used to map atomic coordinates from their values at time 0 to their values at some later time t. In this case, let

$$\vec{r}_i(t) = \sum_{n=1}^{N_\Gamma} \Gamma_n(t) \vec{f}_n(\vec{r}_i^{\,0}) \tag{2}$$

where $\vec{r}_i^{\,0}$ is the initial position vector of atom i, $\vec{r}_i(t)$ is the position vector of atom i at time t, and N_Γ is the number of global coordinates. Thus, all of the time dependence of the atomic position vectors is in the N_Γ's. The transformation functions in Eq. (2) should be chosen such that invariant quantities, such as the linear and angular momenta of an isolated system, are preserved during a transformation.

To minimize the energy of a system, a set of steepest descent equations for the pseudotime evolution of the global coordinates instead of atomic coordinates need to be solved:

$$\frac{d\Gamma_n}{dt} = -q_n' \frac{\partial U}{\partial \Gamma_n} = -q_n' \sum_{i=1}^{N_\Gamma} \frac{\partial U}{\partial \vec{r}_i} * \frac{\partial \vec{r}_i}{\partial \Gamma_n} = -q_n' \sum_{i=1}^{N_\Gamma} \frac{\partial U}{\partial \vec{r}_i} * \vec{f}_n(\vec{r}_i^{\,0}) \quad \text{n=1 , 2, ..., } N_\Gamma \tag{3}$$

The RHS can be computed via the chain rule very efficiently. With global coordinates solved, a reverse transformation from the global coordinates to current atomic coordinates is applied to obtain the time evolution of actual atomic coordinates.

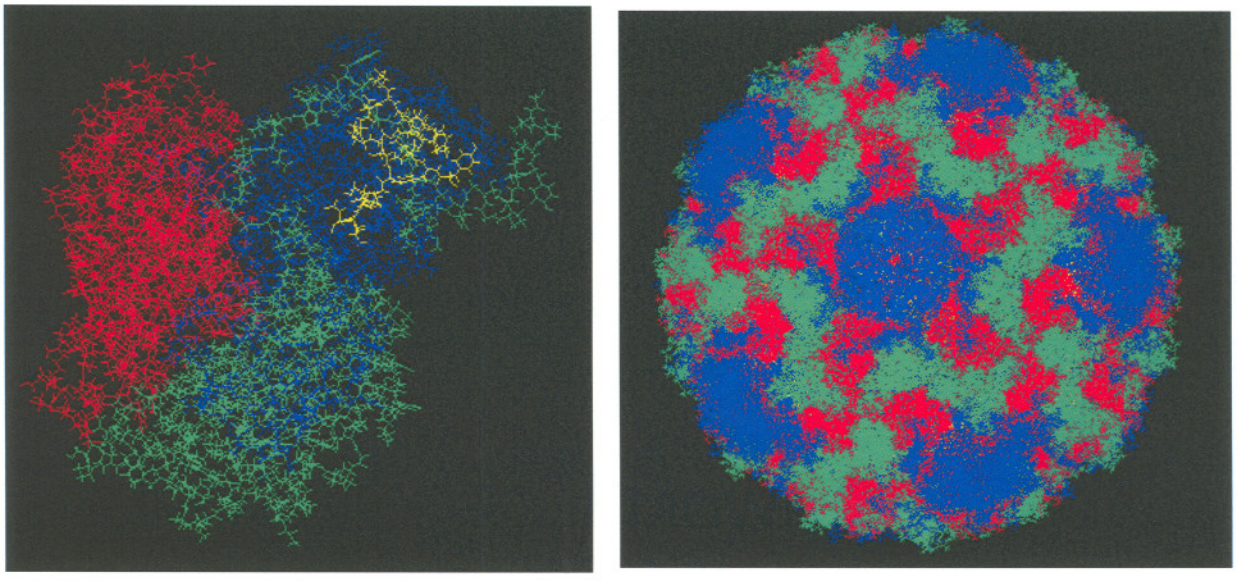
The code of space-warping method has also been finished by *K. Jaqaman et al.* under Fortran 77. It will be transplanted to Fortran 90 and parallelized with OpenMP

in a similar manner as I rewrote the 3D nonlinear PB solver using Fortran 90.

4. Results and Discussion

4.1 Icosahedral Virus Capsid Generator

With IVGP, a complete icosahedral virus capsid with structural data and force field parameters can be generated. The CCMV capsid output in Figure 1 has been shown in Part 2. Another example is the generated HRV capsid in Figure 2. There are 4 protein chains in a HRV protomer (Figure 2(a)). The HRV capsid with icosahedral symmetry is made of such 60 protomers. Pentamers and hexamers can be seen clearly on the capsid surface.



(a) (b)

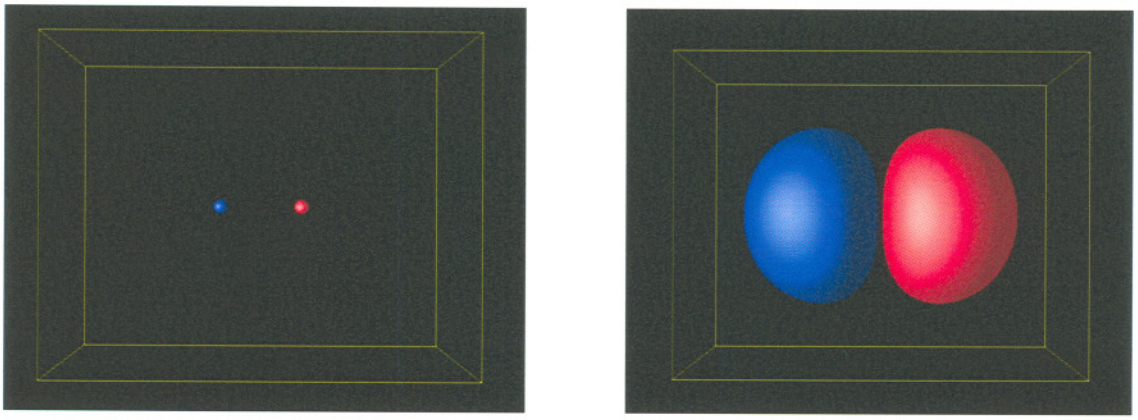
Figure 2 (a) the HRV protomer (A in blue, B in red, C in green and D in yellow) and (b) The HRV capsid (a T=4 crystal lattice for a truncated icosahedron) (Images prepared with

MDL Chime).

4.2 PB calculations

Before studying the conformational changes of complete virus capsids, the consistency of our 3D nonlinear PB model is tested with a two-charge system and the feasibility of integrating this PB model into VirusX to calculate electrostatic forces on virus fragments is investigated. The second part is demonstrated with PB implementation on the protomer and capsid of CCMV and HRV.

First, a theoretical two-charge system is studied. As shown in Figure 3 (a), two theoretical charges of -1e and +1e are immersed in a 1:1 electrolyte and separated at a distance of 10Å. Both their radii are set as 1.0Å. The system size is set as such that the electrostatic potential is cut off at a distance of two Debye lengths away from the charges. The nonlinear PB equation of the system with a resolution of 245 x 195 x 195, 0.2 Å grid spacing is solved with electrolyte ionic strength set as 0.15M. The 3D electrostatic potential isosurfaces of -0.1kT/e (blue) and +0.1kT/e (red) in PB solution are shown in Figure 3(b).



(a) (b)

Figure 3 (a) Two-charge system immersed in a 1:1 electrolyte: Two charges of -1e and +1e are separated at a distance of 10Å and both their radii are set as 1.0Å and (b) 3D electrostatic potential isosurfaces of -0.1kT/e (blue) and +0.1kT/e (red) of the system with a resolution of 245 x 195 x 195, 0.2Å grid spacing (Images prepared with OpenDX).

The electrolyte ionic strength effects on the electrostatic potential profiles and energies of the system are also studied. The ionic strength is changed from 0.05M to 0.25M with a difference of 0.05M. The potential cut-off distances are fixed as two times of the largest Debye lengths (2×13.73Å) of the electrolyte, which is calculated from the lowest ionic strength (0.05M). Calculation results show that the electrostatic energy of the system decreases gradually as the electrolyte ionic strength increasing from 0.05M to 0.25M (Figure 4). It is consistent with Debye-Hückel theory. In this study, the concentrations of mobile ions are changed to increase the electrolyte ionic strength. With higher ionic strength, there will be more counter ions around each charge and more screening. As screening effect increasing, the electrostatic energy of the system will decrease.

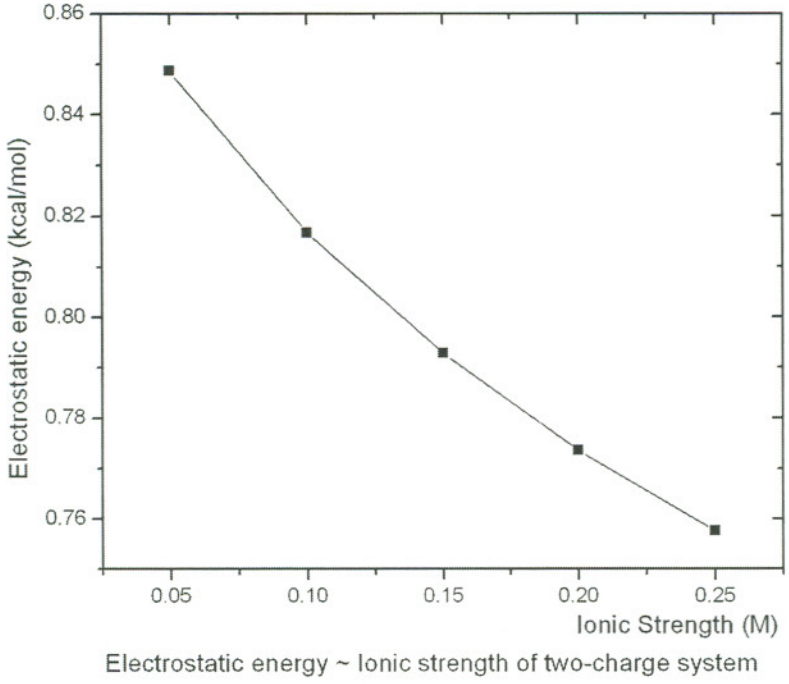
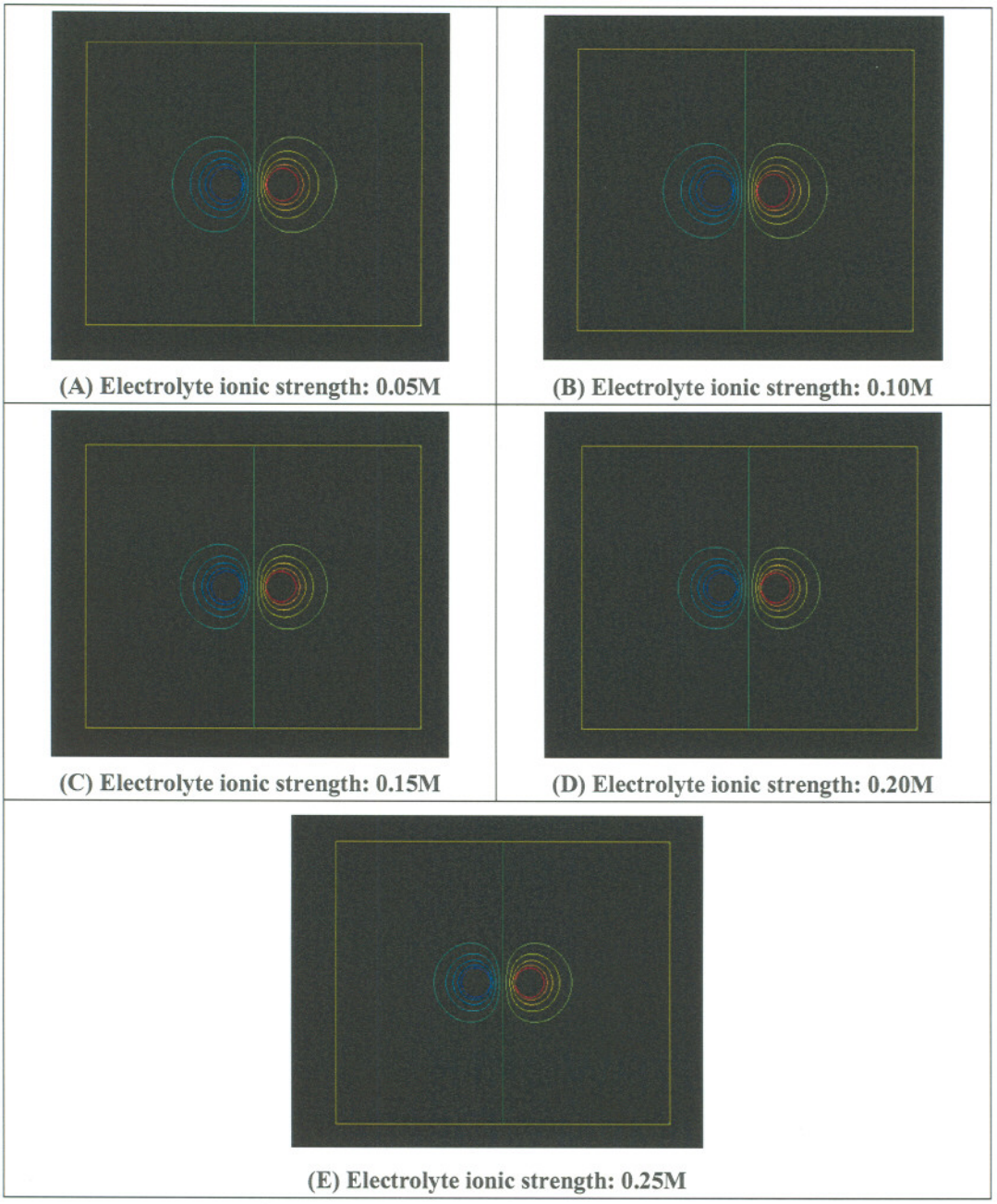


Figure 4 the electrostatic energy of two-charge system decreases as the electrolyte ionic strength increasing from 0.05M to 0.25M.

Along with the electrostatic energies, the electrostatic potential profiles are also obtained in the calculations. Electrostatic potential isosurfaces (from -1.0KT/e (blue) to 1.0KT/e (red) with the same difference of 0.2KT/e) of the two-charge system for electrolyte ionic strength from 0.05M to 0.25M are shown in table 1. As ionic strength increasing, potential isosurfaces become denser. In other words, the potential changes more quickly in the space with higher electrolyte ionic strength, which is consistent with the Debye-Hückel theory. As ionic strength increasing, the Debye length λ_D decreases and the potential changes more quickly in the space according to equation (15) and (16) in Part 3.3.



(A) Electrolyte ionic strength: 0.05M

(B) Electrolyte ionic strength: 0.10M

(C) Electrolyte ionic strength: 0.15M

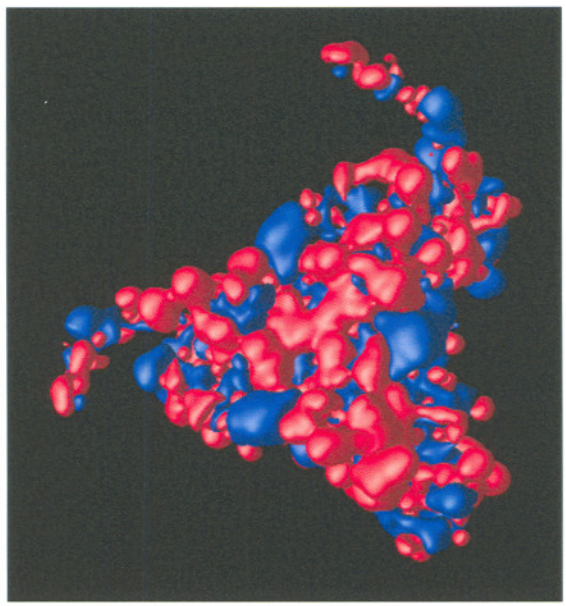
(D) Electrolyte ionic strength: 0.20M

(E) Electrolyte ionic strength: 0.25M

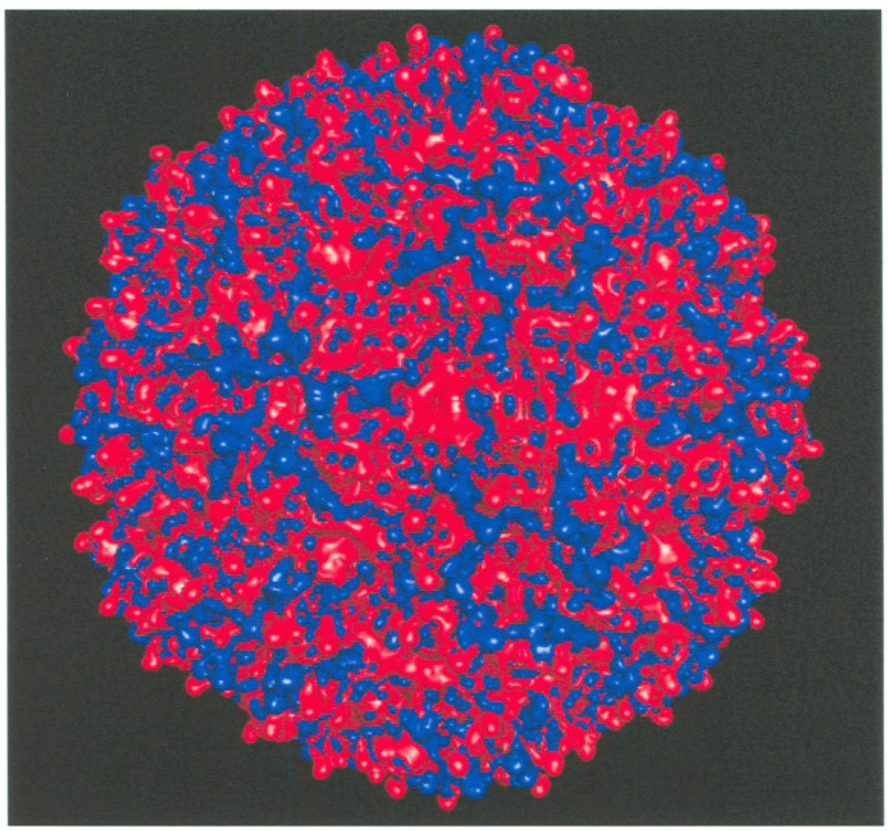
Table 1 Electrostatic potential isosurfaces (from -1.0KT/e (blue) to 1.0KT/e (red) with the same difference of 0.2KT/e) of the two-charge system for different electrolyte ionic strengths (Images prepared with OpenDX).

Second, the PB model is implemented on larger systems, CCMV protomer and capsid. In the simulations, they are all immersed in a 0.15M 1:1 electrolyte medium and the grid spacing are all fixed at 0.5Å. Their 3D electrostatic potential isosurfaces

of -5.0kT/e and +5.0kT/e are shown in Figure 5.



(a)



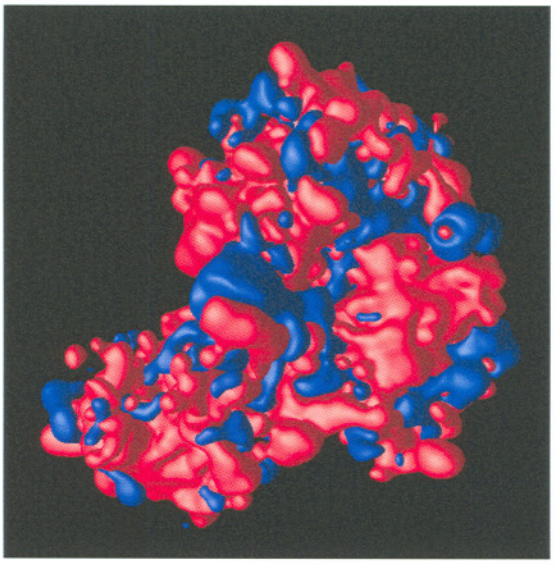
(b)

Figure 5 (a) 3D electrostatic potential isosurfaces of -5.0kT/e (blue) and +5.0kT/e (red) of the CCMV protomer immersed in a 0.15M 1:1 electrolyte with a resolution of 189 x 260 x 246, 0.5Å grid spacing and (b) 3D electrostatic potential isosurfaces of -5.0kT/e (red) and +5.0kT/e (blue) of 432,240 atom CCMV capsid immersed in a 0.15M 1:1 electrolyte with a resolution of 625^3, 0.5Å grid spacing.

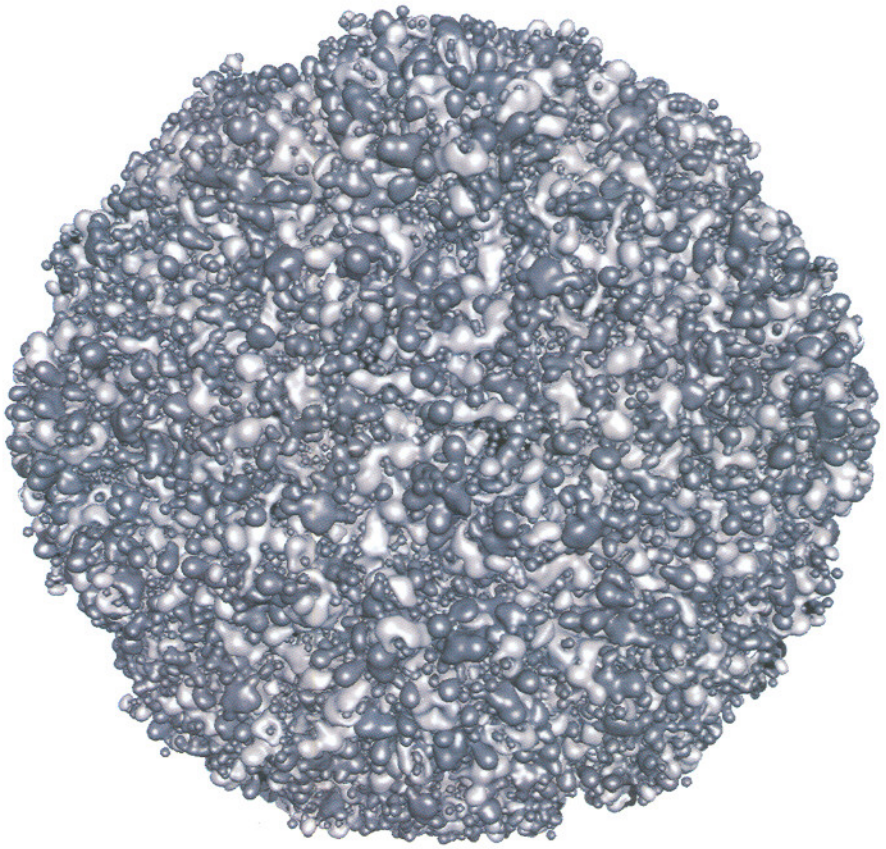
Apart from running PB on CCMV protomer and capsid, similar simulations are implemented on HRV protomer and capsid. The 3D electrostatic potential isosurfaces of -5.0kT/e and +5.0kT/e in their PB solutions are visualized as Figure 6. In both Figrue 5 and Figure 6, the first image (a) was prepared with OpenDX, while for the second one, because the ouput file of PB potential solution is very big (3.2G for CCMV capsid and 4.2G for HRV capsid), we prepared it with our own visualization software tool, Potential Isosurface Renderer (PIR), and it is some different from the first one.

To solve the 3D nonlinear PB equation on large systems, like 432,240 atom CCMV and 739,560 atom HRV capsids, efficient memory usage and fast computation speed are needed. With a resolution of 0.5Å grid spacing, 625^3 and 685^3 grid nodes are required for CCMV and HRV respectively. *A. Sayyed-Ahmad et al.*[3] have shown that our 3D nonlinear PB model requires less memory to solve PB equation than PMG[17, 18] for the same number of unknowns. Obtaining the PB solution on CCMV capsid to an accuracy of 0.01KT/e required approximately 11 hours on 16 processors of IBM SP power4 computers located at Indiana University, and 15 hours for HRV

capsid.



(a)



(b)

Figure 6 (a) 3D electrostatic potential isosurfaces of -5.0kT/e (blue) and +5.0kT/e (red) of the HRV protomer immersed in a 1:1 electrolyte with a resolution of 240 x 241 x 193, 0.5Å grid spacing. (b) 3D electrostatic potential isosurfaces of -5.0kT/e (dark grey) and +5.0kT/e (light grey) of 739,560 atom HRV capsid immersed in a 0.15M 1:1 electrolyte with a resolution of 685^3, 0.5Å grid spacing.

The PB implementations on CCMV and HRV show that it is feasible to integrate our 3D nonlinear PB model into VirusX to compute the electric field of large viral systems. The electrostatic forces on divided fragments of virus capsids can be calculated to replace the 1/r coulomb calculation used in force fields to account for the electric channeling along macromolecules.

5. Conclusion

In this VirusX project, PDB2PQRM, Icosahedral Virus Capsid Generator (IVCG), and the 3D nonlinear PB solver have been accomplished. Calculation results show that it is applicable to integrate our PB solver into VirusX to compute the electric field of large viral systems and electrostatic forces on divided fragments in virus capsids to account for the electric channeling along viral macromolecules.

In our future study, the code for Van der Waals interaction calculation between virus fragments will be developed and integrated with the PB solver into VirusX under the framework of spacing-warping method. OpenMP parallelization scheme

will be implemented in VirusX using Fortran 90.

Acknowledgements

I thank *A. Sayyed-Ahmad* and *K. Tuncay* for developing the original code of 3D nonlinear PB solver under Fortran 77 and *K. Jaqaman et al.* for space-warping method code. Their work provides the basis for this project. I thank *K. Hubbard* for preparing images of 3D potential isofurfaces of CCMV and HRV capsids ((b) in Figure 5 and Figure 6). He tried many times to load big files and make these two images. I thank Professor *P. Ortoleva*, my advisor, for guiding me through this project, giving me important suggestions and helping me with editing this report. I thank all group members at Center for Cell and Virus Theory, Indiana University for giving me help during my research work of this project.

References

1. H. M. Berman et al. The Protein Data Bank. *Nucleic Acids Research* 28, 235-242, 2000.
2. K. Jaqaman, P. J. Ortoleva, New Space Warping Method for the Simulation of Large-Scale Macromolecular Conformational Changes, J. Comp. Chem. 23: 484–491, 2002.
3. A. Sayyed-Ahmad, K. Tuncay, P. J. Ortoleva. Efficient Solution Technique for Solving the Poisson–Boltzmann Equation, *J Comput Chem* 25: 1068–1074, 2004.

4. I. K. Robinson and S.C. Harrison. Structure of the expanded state of TBSV. *Nature* 297: 563–567, 1982.

5. P. K. Sorger, P.G. Stockley and S.C. Harrison. Structure and assembly of turnip crinkle virus II: mechanism of reassembly in vitro. *J. Mol. Biol.* 191: 639–658, 1986.

6. J.A. Speir, S. Munshi, G. Wang, T. S. Baker and J.E. Johnson. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryoelectron microscopy. *Structure 3*: 63–78, 1995.

7. J.E. Johnson and J.A. Speir. Quasi-equivalent viruses: a paradigm for protein assemblies. *J. Mol. Biol.* 269: 665–675, 1997.

8. A. D. MacKerell et al., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, *J. Phys. Chem. B* 102, 3586-3616, 1998.

9. A. D. MacKerell. Empirical Force Fields for Biological Macromolecules: Overview and Issues, *J Comput Chem* 25: 1584–1604, 2004.

10. B. I. Shklovskii. Screening of a macroion by multivalent ions: Correlation-induced inversion of charge. *Physical Review E*, 60(5), 5802-5811, 1999.

11. H. Liu, C. Qu, J.E. Johnson and D.A. Case. Pseudo-atomic models of swollen CCMV from cryoelectron microscopy data. *J. Struct. Biol.* 142: 356-63, 2003.

12. F. Tama and C. L. Brooks III. The mechanism and pathway of this pH induced swelling in cowpea chlorotic mottle virus. *J. Mol. Biol.* 318, 733–747, 2002.

13. D. Zhang et al. Electrostatic interaction between RNA and protein capsid in

cowpea chlorotic mottle virus simulated by a coarse-grain RNA model and a Monte Carlo approach. *Biopolymers*, Vol. 75, 325–337, 2004.

14. R. Isea , C. Aponte , R. Cipriani. Can the RNA of the cowpea chlorotic mottle virus be released through a channel by means of free diffusion? A test in silico. *Biophysical Chemistry* 107, 101–106, 2004.
15. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research* 32, 665-667, 2004.
16. V. S. Reddy et al. Virus Particle Explorer (VIPER), a Website for virus capsid structures and their computational analysis. *J. Virol.* 75:11943-11947, 2001.
17. M. Holst, N. Baker, F. Wang. Adaptive Multilevel Finite Element Solution of the Poisson Boltzmann Equation I. Algorithms and Examples. *Journal of Computational Chemistry*, Vol. 21, No. 15, 1319 1342, 2000.
18. N. Baker, M. Holst, F. Wang. Adaptive Multilevel Finite Element Solution of the Poisson Boltzmann Equation II. Refinement at Solvent-Accessible Surfaces in Biomolecular Systems, Journal of Computational Chemistry, Vol. 21, No. 15, 1343 1352, 2000.